

OPPONENT-COLOR FUSION OF MULTI-SENSOR IMAGERY: VISIBLE, IR AND SAR*

A.M. Waxman, M. Aguilar, R.A. Baxter, D.A. Fay,
D.B. Ireland, J.P. Racamato, and W.D. Ross

Machine Intelligence Technology Group
MIT Lincoln Laboratory
Lexington, MA 02173

ABSTRACT

The primary motivation for multi-sensor image fusion is to combine the complementary information derived from different modality sensors. Building on the work reported in two of our earlier papers from IRIS Passive Sensors 1996, we show how opponent-color processing and center-surround shunting neural networks can be used to develop a variety of image fusion architectures. By emulating single-opponent color processing cells in the retina, and double-opponent color cells in primary visual cortex, we demonstrate an effective strategy for color image fusion as applied to:

- *low-light visible and thermal IR fusion for color night vision,*
- *6-band multispectral fusion for camouflage detection,*
- *EO / IR / SAR multi-modal fusion from separate sensor platforms.*

We have also developed a realtime visible/IR fusion processor from multiple C80 DSP chips using commercially available boards, and use it in conjunction with the Lincoln Lab low-light CCD and an uncooled IR camera. Limited human factors testing of visible/IR fusion has shown improved human performance using our color fused imagery as compared to alternative fusion strategies or either single image modality alone. We conclude that fusion architectures which match opponent-sensor contrasts to human opponent-color pathways will yield fused image products of high image quality and utility.

1. BACKGROUND

The motivations for multi-sensor image fusion go well beyond the desire for multiple views of a scene in order to obtain statistically significant measurements, overcome visual occlusion, or manage a hand-off between sensing modalities. In choosing multiple sensing modalities, one seeks to combine the complementary information obtained from those modalities, and in the case of imaging sensors one can benefit from fusing all image modalities into a single color composite for presentation to the user, or for further processing. Examples addressed in this paper involve the

* Acknowledgements: This work was sponsored by the Air Force Office of Scientific Research and the Defense Advanced Research Projects Agency, under Air Force Contract F19628-95-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and not necessarily endorsed by the U.S. Air Force.

REPORT DOCUMENTATION PAGE

Form Approved OMB No.
0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 01-01-1998	2. REPORT TYPE Conference Proceedings	3. DATES COVERED (FROM - TO) xx-xx-1998 to xx-xx-1998		
4. TITLE AND SUBTITLE Opponent-Color Fusion of Multi-Sensor Imagery: visible, IR and SAR Unclassified		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Waxman, A. M. ; Aguilar, M. ; Baxter, R. A. ; Fay, D. A. ; Ireland, D. B. ;		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME AND ADDRESS Machine Intelligence Technology Group MIT Lincoln Laboratory Lexington, MA02173		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME AND ADDRESS Director, CECOM RDEC Night Vision and electronic Sensors Directorate, Security Team 10221 Burbeck Road Ft. Belvoir, VA22060-5806		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT APUBLIC RELEASE ,				
13. SUPPLEMENTARY NOTES See Also ADM201041, 1998 IRIS Proceedings on CD-ROM.				
14. ABSTRACT The primary motivation for multi-sensor image fusion is to combine the complementary information derived from different modality sensors. Building on the work reported in two of our earlier papers from IRIS Passive Sensors 1996, we show how opponent-color processing and center-surround shunting neural networks can be used to develop a variety of image fusion architectures. By emulating single-opponent color processing cells in the retina, and double-opponent color cells in primary visual cortex, we demonstrate an effective strategy for color image fusion as applied to: ? low-light visible and thermal IR fusion for color night vision, ? 6-band multispectral fusion for camouflage detection, ? EO / IR / SAR multi-modal fusion from separate sensor platforms. We have also developed a realtime visible/IR fusion processor from multiple C80 DSP chips using commercially available boards, and use it in conjunction with the Lincoln Lab low-light CCD and an uncooled IR camera. Limited human factors testing of visible/IR fusion has shown improved human performance using our color fused imagery as compared to alternative fusion strategies or either single image modality alone. We conclude that fusion architectures which match opponent-sensor contrasts to human opponent-color pathways will yield fused image products of high image quality and utility.				
15. SUBJECT TERMS				
16. SECURITY CLASSIFICATION OF: a. REPORT Unclassified		17. LIMITATION OF ABSTRACT Public Release	18. NUMBER OF PAGES 19	19. NAME OF RESPONSIBLE PERSON Fenster, Lynn lfenster@dtic.mil
b. ABSTRACT Unclassified		19b. TELEPHONE NUMBER International Area Code Area Code Telephone Number 703767-9007 DSN 427-9007		
				Standard Form 298 (Rev. 8-98) Prescribed by ANSI Std Z39.18

fusion of visible and LWIR imagery, multispectral fusion of six bands from visible through SWIR, and fusion of visible, IR, and SAR (synthetic aperture radar) imagery. These modalities image very different physical properties of a scene, and their complementary information is well known. In the case of SAR imagery, with its unusual non-literal appearance, it is very useful to render the SAR information in a literal context as provided by the visible image. In the case of targets in the hide, it is clearly more difficult to spoof a multi-sensor system. Image fusion can also be expected to reduce a user's workload and enhance performance (accuracy and speed) with regard to scene comprehension, target pop-out, change detection between modalities, and even reduce operator fatigue. It remains the domain of human factors testing to determine if, indeed, any of these expectations are realized in practice.

At the *IRIS Passive Sensors 1996* meeting, we presented two papers that introduced our approach to multi-sensor and multispectral color image fusion based on the concepts of opponent-color neural networks (Gove, Cunningham & Waxman, 1996; Waxman et al., 1996c). We had already presented related work in the contexts of synthetic vision for vehicle guidance (Waxman et al., 1995a, 1995b), and target detection and recognition (Waxman et al., 1995c). Two patents have also been pursued (Waxman et al., 1996a, 1996b). This paper reports substantial progress in this approach to image fusion, with new applications described.

At *Passive Sensors 1996*, our dual-sensor imaging system utilized an intensified-CCD and an uncooled LWIR camera of lower resolution, with realtime fusion first implemented on an 8-bit integer video-rate computer and later on dual-C80 DSP boards. This provided the first example of visible/IR color fused night vision. Dual-sensor fusion was achieved using a retinal processing architecture consisting of two layers of center-surround receptive fields performing contrast enhancement with adaptive dynamic range compression, and single-opponent color contrast processing, from which three fields would drive a color display. The multispectral fusion system extended the neural architecture beyond the retinal level, to include double-opponent color contrast fields as found in primary visual cortex. We had demonstrated color fusion on three-band IR imagery (SWIR, MWIR, LWIR sub-bands), with substantial improvement of visual quality and detail over that obtained by mapping three imaged bands (either directly or their principal components) to the red, green, and blue channels of a display.

Prior to our introduction of opponent-color fusion strategies, other methods for image fusion were rooted in pixel-level choice or blending of modalities, aimed at maximizing contrast and implemented on multiscale image representations (Burt & Kolczynski, 1993; Ryan & Tinkler, 1995; Toet, 1990, 1992; Toet et al., 1989). The results are grayscale fused images, which don't support target detection as accurately as our color fused images do in human factors tests (Steele & Perconti, 1997; Toet et al., 1997). This motivated Toet, a pioneer in visible/IR image fusion, to develop another color fusion method (Toet & Walraven, 1996). This alternative color fusion method did provide improved target detection results, comparable to our own (on the same imagery with the same human subjects), however, the quality of the imagery was inferior, often resembling colored cartoons (comparisons of the resulting imagery can be found in Toet et al., 1997). Thus, in assessing the utility of fused imagery for select tasks such as target detection and localization, we found that one must be careful to not loose sight of the importance of image quality, for it certainly plays a role in object recognition tasks.

Since the time of *Passive Sensors 1996*, we have deepened our understanding of dual-sensor color fusion (Waxman et al., 1997), made progress in the design of fusion architectures and realtime implementations (Savoye et al., 1996a), obtained encouraging results from human factors experiments with color fused imagery (Steele & Perconti, 1997; Toet et al., 1997; Waxman et al., 1996d), and conducted field demonstrations and comparisons (Waxman et al., 1996e). We have

recently extended the domain of applicability of opponent-color image fusion to six-band multispectral imagery for camouflage detection, and to the fusion of visible (EO), mid-wave infrared (IR), and synthetic aperture radar (SAR) imagery collected on multiple platforms at different times. We will summarize here the principals and architectures we have developed, and show results from various sensor collections. At the *Passive Sensors 1998* conference, results were shown to illustrate realtime color fusion of the new Lincoln Laboratory 640x480 pixel low-light CCD with uncooled LWIR imagery from a Raytheon TI Systems camera. A separate paper reports progress on the development of the Lincoln low-light CCD for night vision applications (Reich et al., 1998; Savoye et al., 1996b).

2. BIOLOGICAL PROCESSING DESIGNS AND PERCEPTUAL MOTIVATIONS

The basis of our computational approach to image fusion derives from biological models of color vision and visible/IR fusion. In the case of color vision in monkeys and man, retinal cone (*i.e.*, detector) sensitivities are spectrally broad and overlapping, but the images are quickly contrast enhanced *within bands* by spatial opponent processing via cone-horizontal-bipolar cell interactions creating both ON and OFF *center-surround* channels (Schiller, 1992). These signals are then color-contrast enhanced *between bands* via interactions among bipolar, sustained amacrine, and single-opponent color ganglion cells (Schiller & Logothetis, 1990; Gouras, 1991), all within the retina. Further color processing in the form of double-opponent color cells is found in the primary visual cortex of primates (and the retinas of some fish). Opponent processing interactions form the basis of such percepts as color opponency, color constancy, and color contrast, though the exact mechanisms are not fully understood (Kaiser & Boynton, 1996). A significant insight that one obtains from these neurological findings, is that nonlinear center-surround receptive fields come in many varieties, are used to process imagery within and between bands, are the substrate for opponent processes in vision, and in general play an enormous role in the hierarchical design of biological image processors.

Examples of cross-modality fusion are also known. Fusion of visible and thermal IR imagery has been observed in several classes of neurons in the optic tectum (the evolutionary progenitor of the primate superior colliculus) of rattlesnakes and pythons (pit vipers and boid snakes, respectively), as described by Newman and Hartline (1981, 1982). These neurons display interactions in which one sensing modality (e.g., IR) can enhance or depress the response to the other sensing modality (e.g., visible) in a strongly nonlinear fashion. These tectum cell responses relate to (and perhaps control) the attentional focus of the snake as observed by its striking behavior. This discovery predates the observation of bimodal visual/auditory fusion cells observed in the primate superior colliculus (King, 1990). Moreover, these visible/IR fusion cells are suggestive of ON and OFF channels feeding single-opponent color-contrast cells; a strategy which forms the basis of our computational models.

Our multi-sensor image fusion architectures are constructed from hierarchies of center-surround receptive fields, with organization mimicking that of the retina and first area (V1) of visual cortex. We restrict ourselves here to primarily the color processing stream, though our inclusion of SAR imagery will benefit from models of form processing (involving V1, V2 and V4). We will make use of many types of center-surround fields, though all are well represented in the primate color vision system. The results are consistently color fused products of high image quality. We have found this to be the case across many different sensor systems.

The advantage of color fused imagery is due to the fact that the user's visual system can exploit this coloring to aid perceptual *pop-out* of extended navigation cues and compact targets (Wolfe et al., 1989; Grossberg et al., 1994). Image coloring has long been known as an aid to

interpretation (Fink, 1976), as it allows for many more color contrasts to separate target from background than does a limited grayscale (Barbur & Forsyth, 1990). Moreover, color displays are known to maintain higher vigilance levels in users than similar grayscale displays (Widdel & Pfendler, 1990). It is common experience (e.g., watching television and movies) that color imagery generates greater and more rapid scene comprehension than grayscale imagery does. We have every reason to expect that color fused imagery of high quality will lead to higher levels of human performance, and the human factors experiments conducted so far have borne this out (Steele & Perconti, 1997; Toet et al., 1997). The ability to generate a rich color percept from dual-band imagery was first demonstrated experimentally in the visible (*red and white imagery*) domain by Land (1959a,b), and motivated his famous *retinex theory* of color vision (Land, 1983) which itself lacked any notion of opponent-color!

3. OPPONENT-PROCESSING NEURAL NETWORK

The computational model that underlies all the opponent processing stages utilized here is the feedforward center-surround shunting neural network of Grossberg (1973, 1988a; also see Elias & Grossberg, 1975). It is used to enhance spatial contrast within individual sensor bands, to adaptively normalize and compress dynamic range through local gain control, to create both positive (ON) and negative (OFF) polarity contrast images, and to create single-opponent color-contrast images between bands or between sensing modalities. It is used to extract cross modality contrast as a form of new information to visualize.

The large variety of center-surround receptive fields utilized in our fusion architectures are based on the same opponent-processing neural network, whose dynamics (and equilibrium) are described at each pixel ij by the equations,

$$\frac{dE_{ij}}{dt} = -AE_{ij} + (1-E_{ij})[CI^C]_{ij} - (1+E_{ij})[G_S * I^S]_{ij} \quad (1a)$$

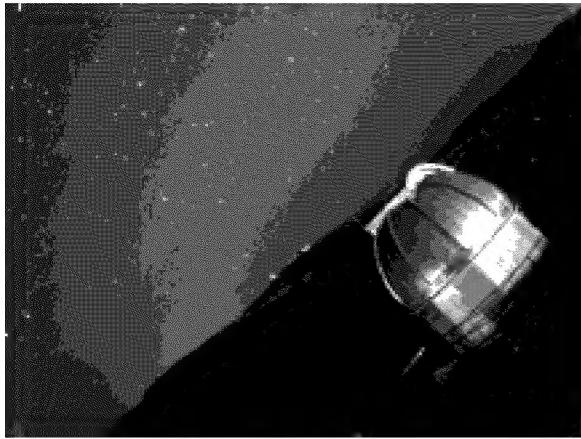
$$E_{ij} = \frac{[CI^C - G_S * I^S]_{ij}}{A + [CI^C + G_S * I^S]_{ij}} \quad (1b)$$

where E is the opponent processed *enhanced* image, CI is the *input* image that excites the single pixel *center* of the receptive field (a single pixel center is used to preserve resolution of the processed images), and I^S is the *input* image that inhibits the gaussian *surround* G_S of the receptive field. Equation (1a) describes the temporal dynamics of a charging neural membrane (*cp. capacitor*) which leaks charge at rate A , and has excitatory and inhibitory input ion currents determined by Ohm's law (the shunting coefficients $(1\pm E)$ act as potential differences across the membrane, and the input image signals modulate the ion selective membrane conductances). Equation (1b) describes the equilibrium that is rapidly established at each pixel (*i.e.*, at frame rate), and defines a type of nonlinear image processing with parameters A , C , and size of the gaussian surround. The shunting coefficients of equation (1a) clearly imply that the dynamic range of the enhanced image E is

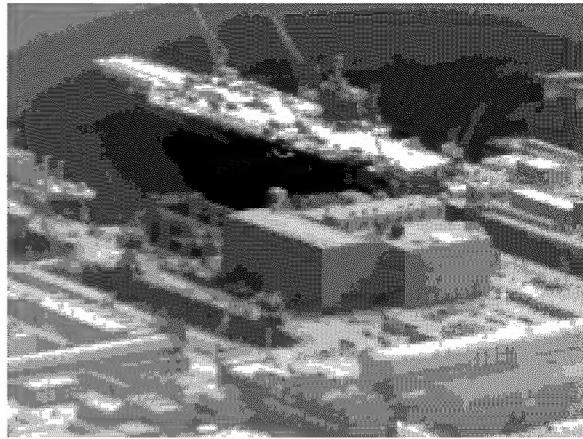
bounded, $-1 < E < 1$, regardless of the dynamic range of the input imagery. When the imagery which feeds the center and surround is taken from the same input image, the numerator of equation (1b) is the familiar difference-of-gaussians filtering which, for $C > 1$, acts to boost high spatial frequencies superimposed on the background. The denominator of equation (1b) acts to adaptively normalize this contrast enhanced imagery based on the local mean in the neighborhood surround. In fact, (1b) displays a smooth transition between linear filtering (when A exceeds the local mean brightness, such as in dark regions) and ratio processing (when A can be neglected as in bright regions of the imagery). This is particularly useful for processing the wide dynamic range imagery obtained with low-light CCDs, FLIRs, and SARs. Following the processing of each image by equation (1b), we remap the resulting bounded dynamic range to an 8-bit integer range by application of a sigmoidal nonlinearity that adapts to the statistics of the processed image. These enhanced images are reminiscent of the *lightness images* postulated in Land's (1983) retinex theory (also see Grossberg, 1988b, on *discounting the illuminant*).

A modified version of equation (1), with an *inhibitory center* and *excitatory surround* is used to create an enhanced OFF image (e.g., a reverse polarity enhanced IR image). With imagery of separate bands/sensors first enhanced and normalized by application of equation (1b), we make repeated use of this opponent processing to form band/sensor-contrast images by feeding one band/sensor to the center and another to the surround region of the model. This results in terms involving band differences and band ratios, both measures used in the remote sensing community. We have found that a surround neighborhood of 7×7 pixels provides high quality results, though our realtime implementations have generally utilized only 3×3 neighborhoods. This simple module forms the computational building block of all our fusion architectures. It involves approximately 35 integer operations per pixel on 16 bit data. We believe it could be realized in an efficient, compact, and low-power implementation on two FPGAs or one ASIC.

Figures 1a-f illustrate the usefulness of this opponent-processing network on single modality imagery obtained from wide dynamic range visible (Fig. 1a-c) and mid-wave IR (Fig. 1d-f) cameras. Figures 1a-c show a 640×480 image taken with a Lincoln CCD camera under starlight conditions. The CCD itself has an enormous dynamic range, of which only a portion is digitized to 12 bits. The bright end of the dynamic range captures the information shown in Figure 1a, while the dark end of the dynamic range is shown in Figure 1b. This 12-bit dynamic range cannot be presented on a typical 8-bit display, however, the adaptively processed image of



1a. Bright end of 12-bit dynamic range.



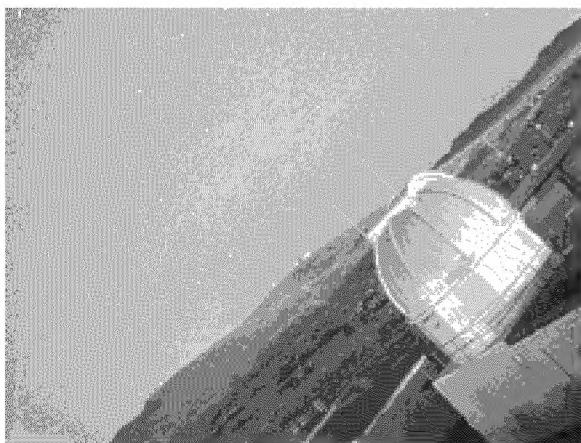
1d. Histogram remapped 12-bit dynamic range.



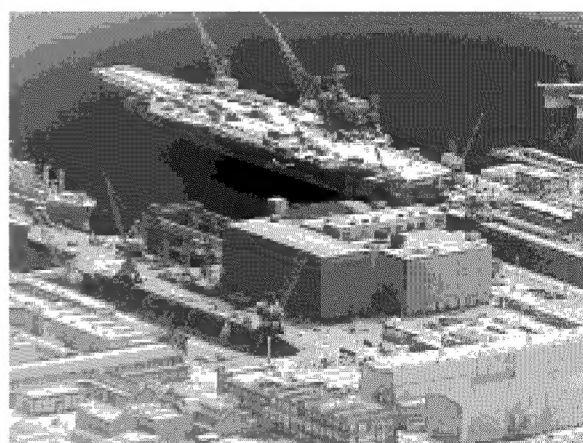
1b. Dark end of 12-bit dynamic range.



1e. Adaptively processed; one iteration.



1c. Adaptively processed full dynamic range.



1f. Adaptively processed; two iterations.

Figure 1. Adaptive processing of wide dynamic range imagery using center-surround shunting neural networks. (a-c) Visible image taken under starlight conditions with a Lincoln Lab CCD camera. (d-f) MWIR image taken during daylight with a Hughes InSb focal plane camera.

Figure 1c has compressed all the contrast information into an 8-bit dynamic range, well suited for display. There are certainly other approaches to compressing dynamic range, besides the adaptive opponent-processing network we utilize. However, most other methods try to remap brightness directly, by manipulating a histogram without regard to spatial distribution of brightness, *i.e.*, contrast. Figures 1d-f illustrate this comparison on a 640x480 MWIR image with 12-bit dynamic range taken with a Hughes InSb focal plane camera flown in a P-3 aircraft. Figure 1d (provided to us by S. Campana of the former NAWC, Warminster) displays a log-brightness histogram equalized version of the data, and shows large areas of saturation at both ends of the dynamic range. Figure 1e is the result of our adaptive processing on the same original data, and displays greater detail, improved sharpness, and far less saturation. And since the adaptive processing is nonlinear, we can iterate the process to reveal still greater detail, as shown in Figure 1f (which was derived by processing Figure 1e). With each iteration of the adaptive processing, more contrast information is being allocated to the fixed dynamic range of the 8-bit grayscale display.

4. VISIBLE/IR FUSION ARCHITECTURES

For night operations (both military and civilian), the reflected visible moon/star-light and the thermally emitted IR bands provide complementary information, and both modalities are routinely used by rotorcraft pilots (currently in the form of helmet-mounted image intensifier tubes and turret-mounted FLIRs). For ground navigation as well as targeting, both modalities are also utilized. A realtime system with low latency, which can display fused imagery to the user is desirable and expected to enhance human performance.

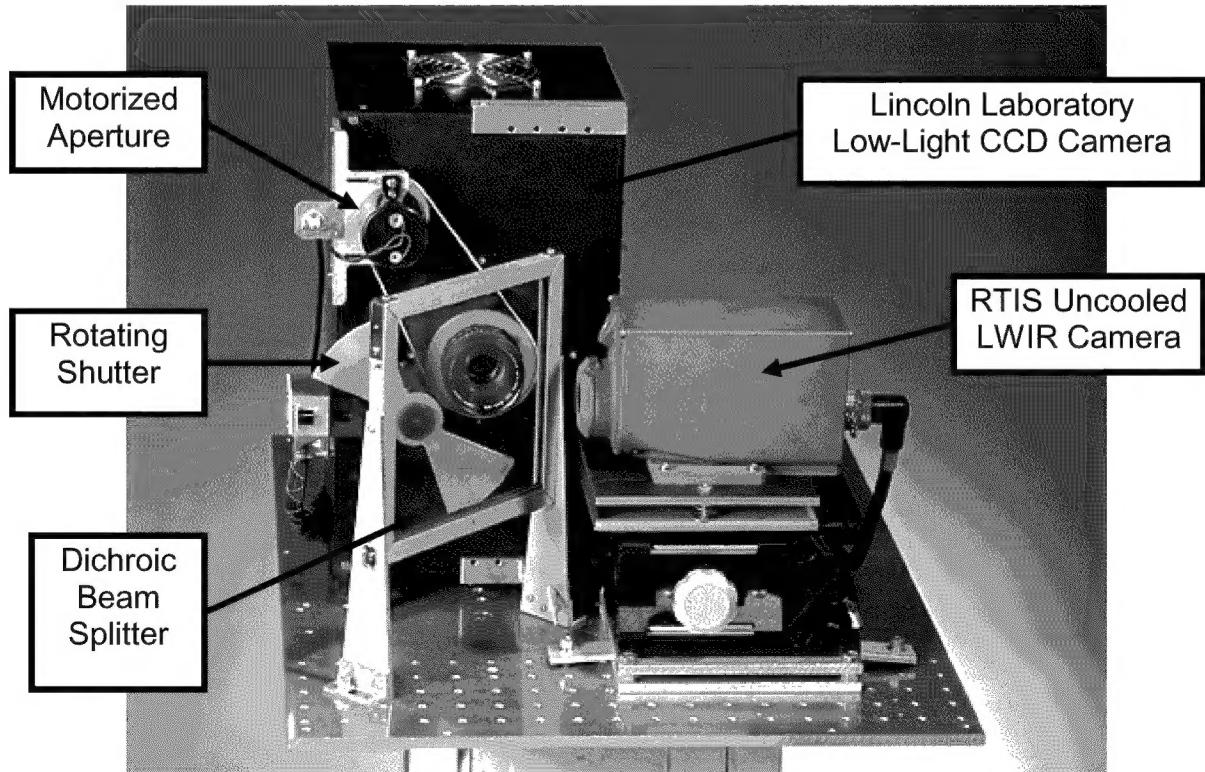


Figure 2. Dual-sensor imaging pod developed at Lincoln Laboratory. It consists of a Lincoln 640x480 pixel low-light CCD, a Raytheon TI Systems 320x240 pixel uncooled LWIR camera, and a dichroic beam splitter.

Figure 2 illustrates a dual-sensor visible/LWIR imaging pod recently constructed at Lincoln Laboratory for the DARPA Integrated Imaging Sensors program. It consists of a Lincoln Lab low-light CCD imager of 640x480 pixel resolution, able to provide useful imagery at 30 frames/sec (or slower) below starlight illumination levels (Reich et al., 1998; Savoye et al., 1996b), an uncooled ferroelectric LWIR thermal imager of 320x240 resolution from Raytheon TI Systems (Flannery & Miller, 1992), and a dichroic beam splitter that transmits the visible-NIR band but reflects the LWIR band. The low-light CCD imager itself has a dynamic range of five orders-of-magnitude, though only three orders-of-magnitude are currently supported by the 12-bit A/D converter. The uncooled LWIR camera outputs a standard analog video stream which is then digitized to 8-bit pixels. We will soon acquire a Lockheed Martin uncooled microbolometer LWIR camera which outputs 15-bit digital imagery. The lenses utilized on both cameras, in conjunction with the beam splitter, provide a nearly registered 40° field of view. Deviations from registration (magnification and distortion) are compensated for in the realtime fusion processor.

The neural architecture utilized to fuse visible/LWIR imagery obtained with sensors of unmatched resolution and quality, such as those of Figure 2, is constructed from center-surround opponent processing fields as illustrated in Figure 3, and was presented at the *Passive Sensors 1996* conference (Waxman et al., 1996c).

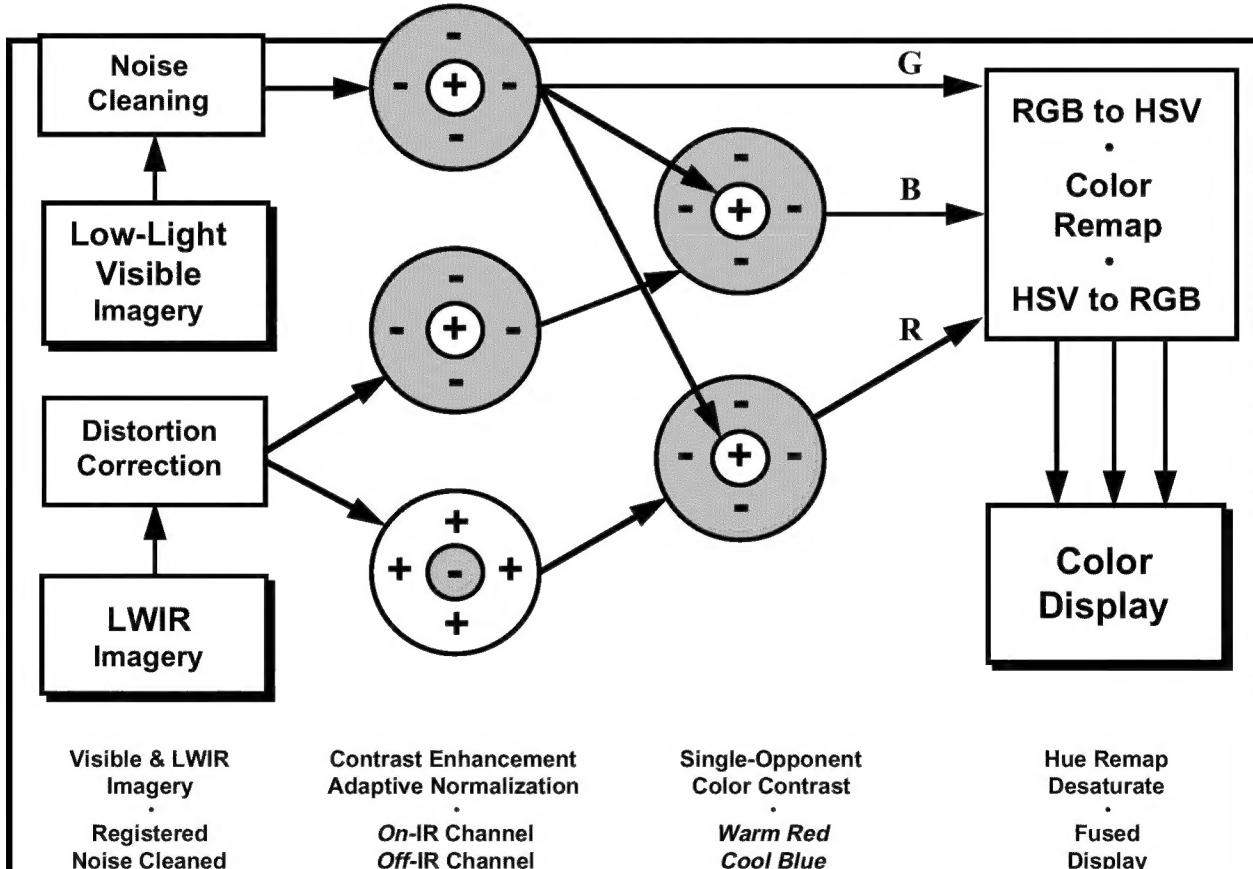


Figure 3. Single-opponent visible/LWIR image fusion architecture built from adaptive center-surround receptive fields. This architecture is well suited to sensors of non-equal resolution, with the higher resolution visible imagery providing input to the centers of the color contrast fields.

Following noise cleaning of the visible imagery (we have explored both realtime median filtering and non-realtime *Boundary Contour/Feature Contour System* processing [Grossberg, 1988b; Waxman et al., 1995c]), and distortion correction to ensure image registration, we form two *grayscale fused* single-opponent color-contrast images using equation (1b) with the enhanced Visible feeding the excitatory centers and the enhanced IR (ON-IR and OFF-IR, respectively) feeding the inhibitory surrounds. In analogy to the primate opponent-color cells (Gouras, 1991), we label these two single-opponent images $+Vis-IR$ and $+Vis+IR$. In all cases, we retain only positive responses for these various contrast images.

Our two single-opponent color/sensor contrast images are analogous to the *IR-depressed-visual* and *IR-enhanced-visual* cells, respectively, of the rattlesnake (Newman & Hartline, 1981, 1982); they even display similar nonlinear behavior. In fact, with the IR image being of lower resolution than the visible image (in the snake, and for man-made uncooled IR imagers), a single large IR pixel may be considered as a small *surround* for its corresponding visible pixel. In this context, our opponent-color contrast images can be interpreted as *coordinate rotations* in the color space of *Visible vs. IR*, along with *local adaptive scalings* of the new color axes. Such color space transformations were fundamental to Land's (1959a,b, 1983) analyses of his dual-band *red and white* colorful imagery.

To achieve a natural color presentation of these opponent images (each being an 8-bit grayscale image), we assign the following color channels (8-bits each) to our digital imagery: (1) enhanced *Vis* to *Green*, (2) $+Vis-IR$ to *Blue*, and (3) $+Vis+IR$ to *Red*. These channels are consistent with our natural associations of *warm red* and *cool blue*. Finally, as shown in the architecture of Figure 3, these three channels can be interpreted as *R,G,B* inputs to a color remapping stage in which, following conversion to *H,S,V* (hue, saturation, value) color space, hues can be remapped to alternative “more natural” hues, colors can be desaturated, and then reconverted to *R,G,B* signals to drive a color display. The result is a high quality *fused color* presentation of visible/IR imagery.

For the case of comparable resolution visible and LWIR sensors, as is obtained with an intensified-CCD and a cryogenically cooled scanning FLIR (or alternatively, the Lincoln Lab low-light CCD and the Hughes InSb MWIR imager, both of 640x480 resolution), we utilize the more symmetric fusion architecture shown in Figure 4, also constructed from center-surround opponent processing fields. Notice that a broadband brightness channel is also formed which can drive a grayscale display for comparison of gray fused to color fused imagery (the subject of some debate, to be reconciled by human factors testing). Again, three output channels can drive a color display directly, or can feed a color remap module which operates in *H,S,V* color space.

Figures 5a-f illustrate visible/LWIR fused results from both of these architectures. Figures 5a,b show enhanced visible imagery (obtained with an intensified-CCD) and enhanced thermal IR imagery (obtained with an uncooled LWIR camera from T.I.), collected with Lincoln's early generation dual-sensor pod under conditions of dusk. The detail on the ground is revealed in the visible, while the horizon and water-line are revealed in the low-resolution LWIR. The fused color image shown in Figure 5c is obtained using the architecture shown in Figure 3, without any remapping of colors before display. Notice how this fused result combines the complementary information provided by the source imagery, and how natural and detailed it appears.

Figures 5d-f illustrate the case of comparable resolution sensors, an intensified-CCD and a 1st-generation FLIR, using the architecture of Figure 4, and including a standard color remap for operations over ground. The source imagery, shown in Figures 5d,e, was obtained during nap-of-the-earth helicopter flight at night, and was provided to us by the Army NVESD. Both digital stills and videotapes for realtime fusion were used to support a human factors study at NVESD which compared the effectiveness of several alternative fusion algorithms including our own (Steele &

Perconti, 1997). Realtime fusion of two analog video streams was implemented on two commercial C80 boards (from Ariel Corp.) in a desktop PC. The fused image, shown in Figure 5f clearly illustrates the combining of complementary information. This time the horizon, tree detail and cast shadows (under quarter-moon), and saturated bloom due to a tower light comes from the intensified-CCD image, while the road is detected in the black-hot FLIR image. The color fused result provides high image quality, and supports enhanced depth perception down the road.

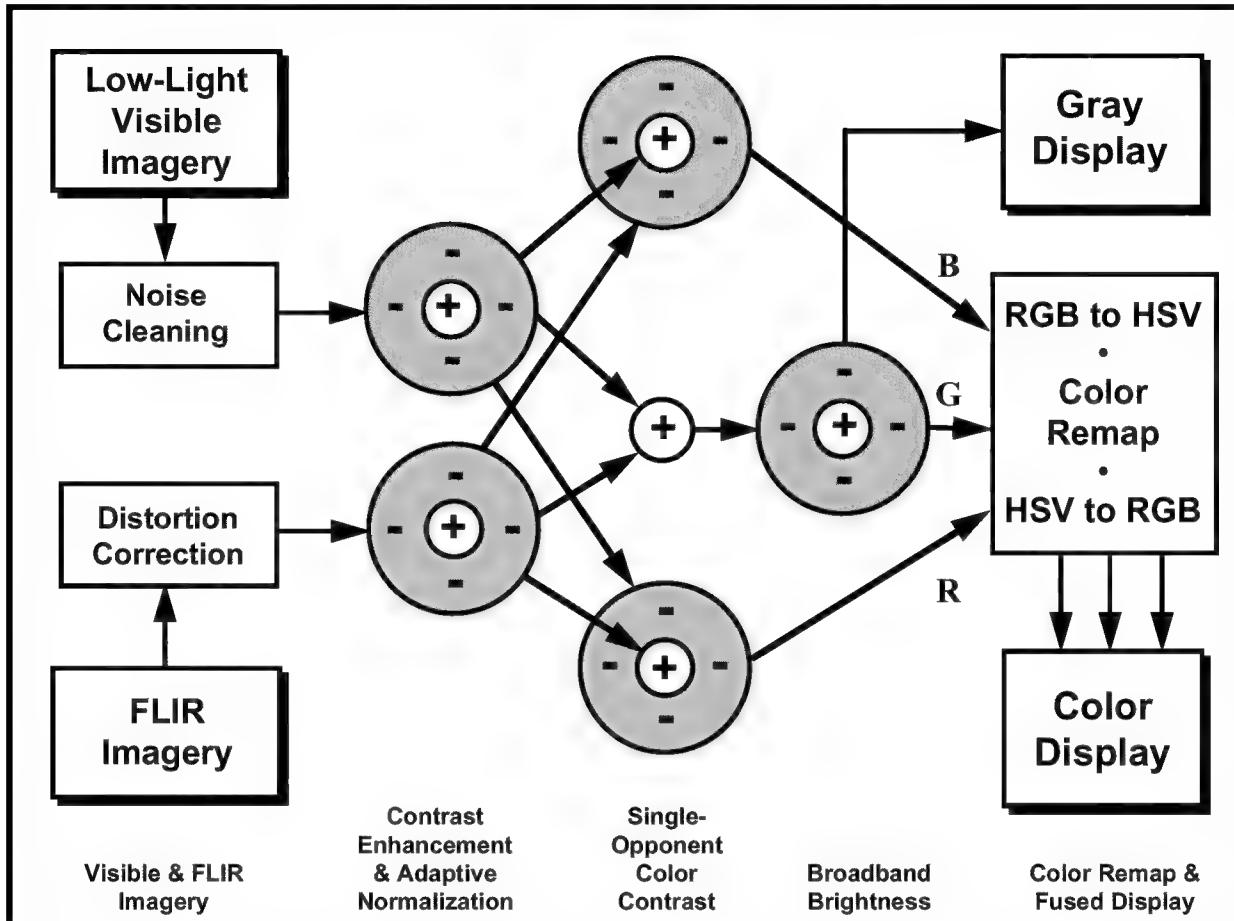


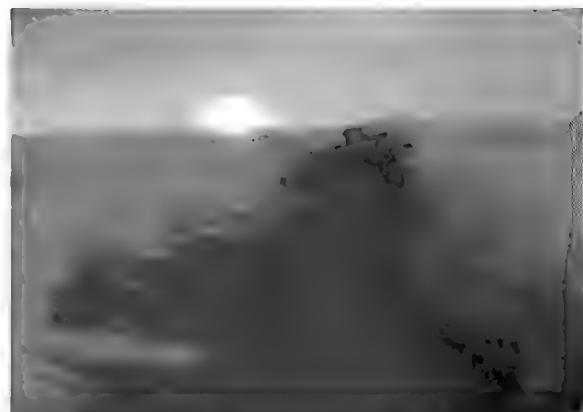
Figure 4. A symmetric, single-opponent visible/FLIR image fusion architecture well suited to sensors of comparable resolution. Both gray and color fused images are created in realtime.

An example of smokescreen penetration, obtained with the fusion architecture of Figure 4, is shown in Figures 6a-c. This imagery was collected by the Canadian military for a NATO study group on image fusion, and was kindly provided to us by A. Toet of TNO, the Netherlands. Though collected during the day, it clearly shows how scenic context can be captured in the visible, while hot targets are easily detected through smoke by a FLIR. However, neither image alone nor the two together convey the quality of information provided by the color fused result. In Figure 6c we can clearly see a tow-truck in front of the smoke, a helicopter behind the smoke, and men and equipment in the smokescreen itself. We have processed a sequence of 100 frames from this collection, and the fused results clearly portray the moving vehicles and men.

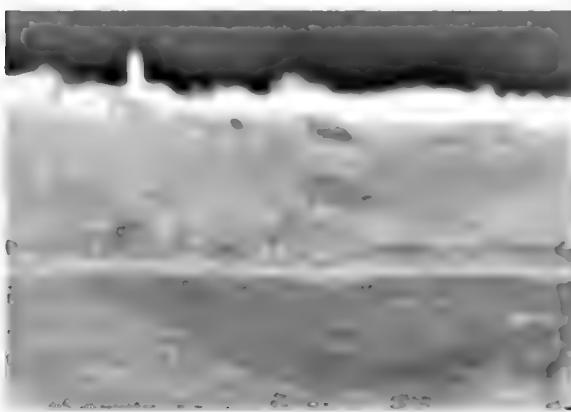
We have recently developed a realtime visible/IR color fusion processor to support wide dynamic range digital imagery provided by the Lincoln Lab low-light CCD and an uncooled IR camera (either the Raytheon TI Systems analog video camera or the Lockheed Martin IR Systems 15-bit digital camera, or for that matter a cryo-cooled InSb camera). We utilize a set of four Matrox Genesis C80 boards, providing for dual-digital video input and six C80 processing nodes, in an industrial PC rack-mount chassis, with a Pentium host processor card. The fusion processor measures 19"x18"x9", weighs 41 lbs., and consumes under 150 watts. With a dedicated ASIC implementation of the center-surround processing of equation (1b), a fusion architecture such as Figure 4 could be realized on a single board.



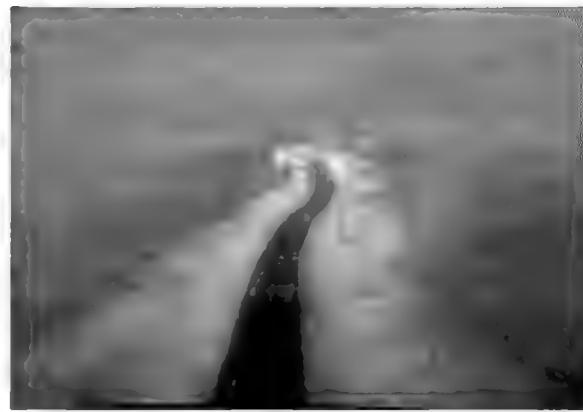
5a. Enhanced IICCD visible image.



5d. Original IICCD visible image.



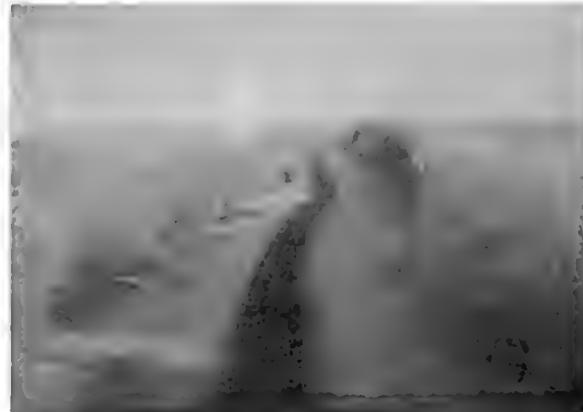
5b. Enhanced LWIR (uncooled) image.



5e. Original thermal IR (FLIR) image.



5c. Fused result derived from 5a,b.



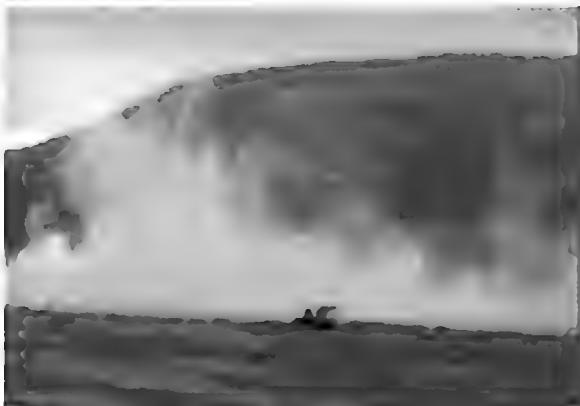
5f. Fused result derived from 5d,e.

Figure 5. Color night vision by fusion of image-intensified visible and thermal IR imagery.

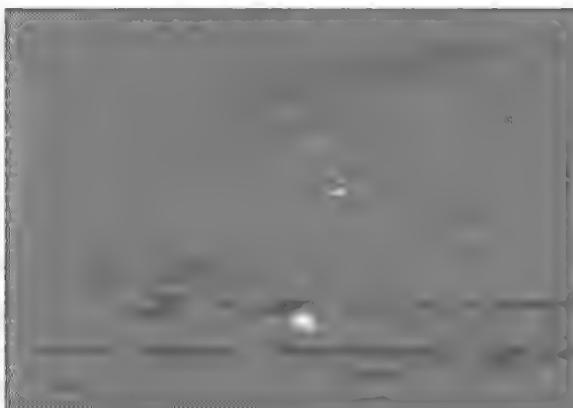
(a-c) Lincoln imagery using sensors of different resolutions; fused using the architecture in Fig 3.

(d-f) Army NVESD helicopter imagery using sensors of similar resolution; fused using the architecture in Fig 4.

Approved for public release;
distribution is unlimited.



6a. Original visible (CCD) image.



6b. Original thermal IR (FLIR) image.



6c. Fused result derived from 6a,b.

Figure 6. Smokescreen penetration by color fusion of visible and thermal IR imagery. Canadian DREV imagery of similar resolution; fused using the architecture in Fig. 4.



8a. Enhanced visible (R, G, B) image.



8b. Fused result derived from 6 bands.

Figure 8. Camouflaged target pop-out by color fusion of multispectral imagery (R, G, B, NIR, SWIR1, SWIR2 bands). Fusion using the architecture in Fig. 7. Targets appear as blue objects.

5. SIX-BAND VISIBLE / NIR / SWIR FUSION ARCHITECTURE

When fusing imagery from more than two bands/sensors, it is necessary to go beyond the single-opponent “retinal” architectures of Figures 3&4. At *Passive Sensors 1996* we introduced a double-opponent “cortical” architecture for the fusion of 3-band IR imagery (Gove, Cunningham & Waxman, 1996). That architecture has been simplified somewhat here, and adapted for 6-band imagery (red, green, blue, near-IR, and 2 short-wave IR bands) collected with an ERIM sensor. All bands are collected in a spatially registered format, though they occupy very different parts of the dynamic range. The architecture for color fusion of these six bands is shown in Figure 7.

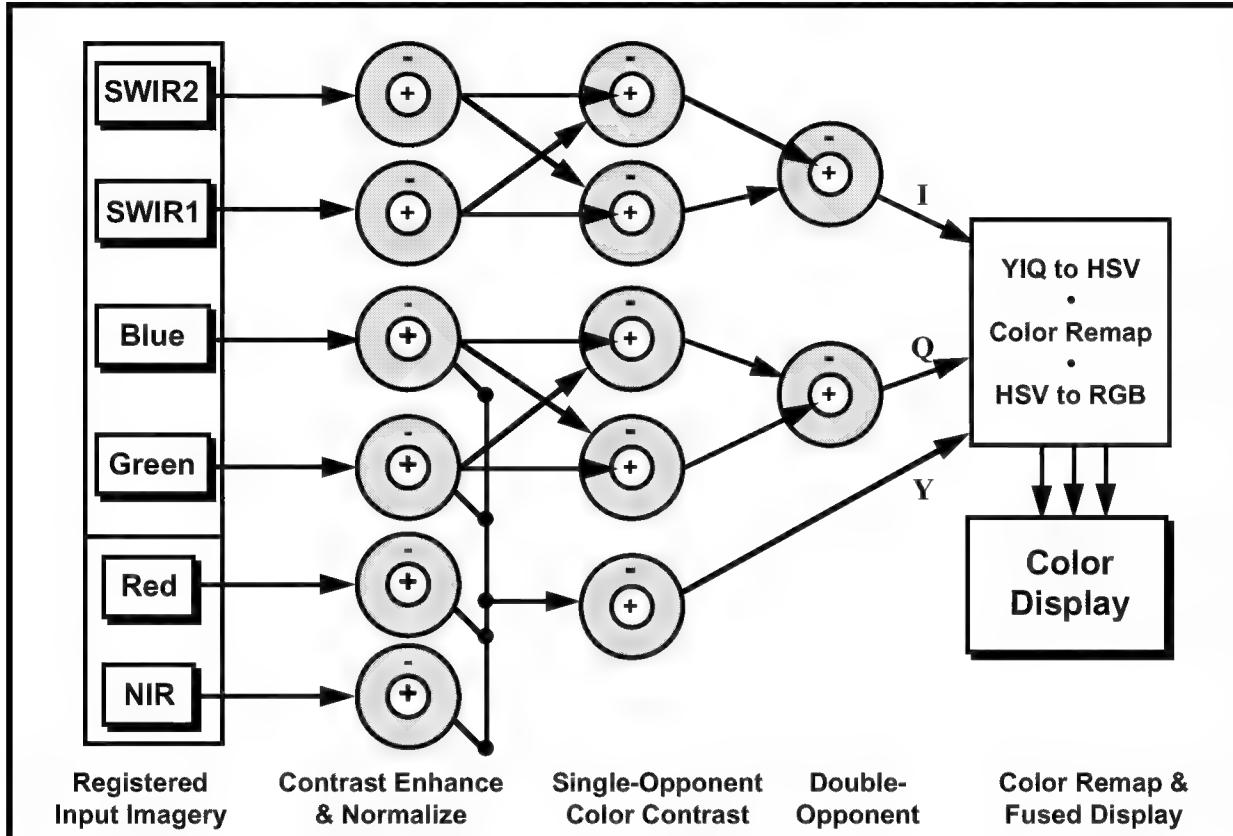


Figure 7. Double-opponent architecture for fusing six bands of imagery into one color product.

This double-opponent architecture is well suited for fusing three or four image bands. In this case we have selected green, blue, and the two SWIR bands as the primary bands, with red and NIR contributing only to the brightness. This particular combination of bands was chosen to illustrate the detection of camouflage, which is designed to match foliage in the visible and NIR bands, but which deviates from foliage in the SWIR bands. In fact, it is the contrast between the two SWIR bands in Figure 7 that is responsible for detecting camouflage. This will become clear when we consider the example in Figure 8.

The architecture in Figure 7 begins with opponent-processing of each band separately in order to enhance contrast and adaptively normalize all the bands. The architecture then organizes into three channels, two opponent-color channels and one broadband channel. The enhanced Red, Green, Blue, and NIR bands are linearly combined to form a single broadband channel which is adaptively processed to enhance brightness contrast and then mapped to a display brightness channel Y . The two opponent-color channels operate on pairs of bands, forming symmetric single-opponent

color-contrasts which are combined into double-opponent contrasts. These double-opponent outputs are then mapped to opponent-color display channels I and Q . The Y,I,Q color space is a very useful mapping, as it attempts to represent human opponent-color channels (and is the basis for NTSC encoding of color television signals). We utilize a simple definition of Y as brightness, I as red vs. green opponency, and Q as blue vs. yellow opponency. In terms of R,G,B display values, we have $Y=[R+G+B]/3$, $I=[R-G]$, $Q=[B-(R+G)/2]$. Thus, the double-opponent outputs and the broadband output can drive a color display directly, or can first undergo a color remapping in the H,S,V color space before display. A color remap typically rotates the hue circle (like the tint control on a television) so that foliage in a scene appears green.

Figure 8 illustrates the utility of six-band fusion to detect camouflage. In figure 8a we show an enhanced visible three-band color image in which the red, green, and blue sensor bands are first adaptively processed to enhance contrast and normalize, and then mapped directly to R,G,B for display. The adaptive processing within bands greatly improves the visibility of detail over the original imagery, but it is still difficult to find the camouflaged targets. In figure 8b we show the double-opponent color fused image derived from six sensor bands. The camouflaged targets now pop out of the scene as three blue objects lined up on the road between the trees on the right side. There is even a decoy object in line with the three targets which does not appear blue in the color fused result. Looking back to figure 8a in the visible bands, we see how the decoy is easily confused with the targets, all of which resemble foliage. Referring back to the architecture in Figure 7, we expect the foliage (being more green than blue) to have a relatively high Q value, and the camouflage (which has stronger SWIR2 than SWIR1 reflectivity) to have a relatively high I value. Thus, before color remapping, the foliage would display as bluish while the targets would be reddish. By rotating the hue circle so as to remap the foliage to green, the resulting remap renders the camouflaged targets as blue. Regardless of the exact choice of color remap, the fused color image provides a large color contrast between the targets and the foliage which makes them visually pop out from the scene.

6. EO / IR / SAR MULTI-PLATFORM FUSION ARCHITECTURE

Three complementary imaging modalities which are commonly used for surveillance are visible (EO), infrared (IR), and synthetic aperture radar (SAR). At these different wavelengths, the imagery reveals very different qualitative information about the scene. Each modality has advantages (and disadvantages), EO being literal, IR (MWIR or LWIR) able to image at night and to reveal thermal structure, and SAR able to operate during the day or night and through clouds, and to reveal metallic objects which act as strong reflectors in the scene. Taken alone, SAR imagery can be very difficult to interpret without a lot of training, and even then it takes more time than a literal EO image. In this era of UAVs carrying multiple imagers of different modalities, it is desirable to be able to fuse the information derived from separate modalities, and delivered by separate platforms, into a single color image. This can be accomplished by first registering the imagery to a common reference frame and then utilizing a modified double-opponent color fusion architecture.

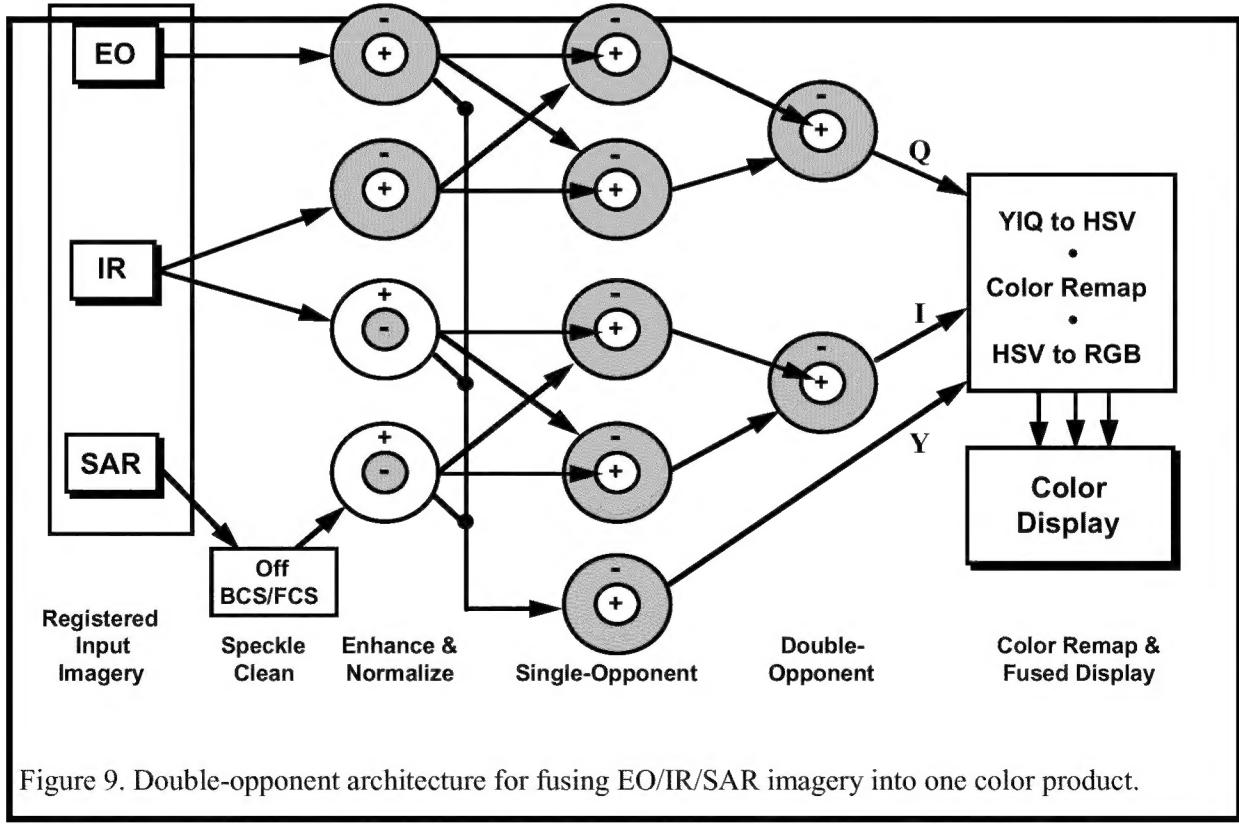
The need to register imagery obtained from different sensor platforms can be considered a separate challenge from that of the multi-sensor image fusion problem. It is straightforward to register a local ground plane across all modalities if the sensor pointing information is known or estimated. However, very different imaging geometries will suffer occlusion and layover problems that cannot be resolved from a single image of each modality. We are currently developing new methods to enable registration and fusion of imagery from multiple sensors.

The double-opponent architecture we utilize for EO/IR/SAR fusion is shown in Figure 9. As was the case with dual-band visible/IR fusion, we utilize both polarities of IR obtained from ON and

OFF opponent-processing fields. The *ON-IR* channel is paired with the *EO* channel to form symmetric single-opponent pairs and a double-opponent output which is mapped to human opponent channel Q . The *OFF-IR* channel is paired with the *SAR* channel to form single-opponent pairs and a double-opponent output which is mapped to human opponent channel I . The brightness channel Y is driven by a linear combination of *EO*, *OFF-IR*, and *SAR*.

Notice that the *SAR* channel is first processed by an *OFF*-polarity *BCS/FCS* network in order to reduce the speckle associated with a typical *SAR* image. The *Boundary Contour System* (*BCS*) and *Feature Contour System* (*FCS*) are neural models of form and shading developed by Grossberg and Mingolla (see Grossberg, 1988b), involving visual processing at the retinal and cortical (*V1*, *V2*, *V4*) levels. *BCS/FCS* processing can be viewed as an elaborate form of boundary completion and edge preserving smoothing, meant to suppress noise without blurring edges or other high contrast features. A simplified version of *BCS/FCS*, involving multi-scale boundary computations, was developed for *SAR* imagery (Waxman et al., 1995c). The first stage of *BCS/FCS* processing is the center-surround field of equation (1b), and an *OFF*-field can be used there to create an *OFF-BCS/FCS* network. This polarity reversal leads to improved speckle reduction, and is compensated for by a following polarity reversal as shown in Figure 9.

The multi-sensor image fusion architecture used here attempts to match *opponent-sensor* channels to *human opponent-color* channels. It also allows for a final remapping of colors in the *H,S,V* color space. It is very effective in rendering the fused scene in a literal manner, while displaying each modality in the context of the others. Simple mappings of the *EO/IR/SAR* imagery (or their principal components) directly to *R,G,B* display channels are completely ineffective, providing poor image quality and having a non-literal appearance.



7. SUMMARY

We have shown that an effective strategy for the fusion of imagery derived from multiple sensors is to emulate the various stages of opponent-color processing in the human visual system. Single-opponent color architectures are sufficient for fusing two sensors, such as a CCD camera and a thermal IR imager. A realtime fusion processor has been developed from commercial DSP boards for fusing the Lincoln Lab 640x480 low-light CCD with an uncooled 320x240 LWIR camera to provide color night vision. Double-opponent color architectures are effective for fusing three or four image bands, and we have also demonstrated the fusion of six bands of visible/NIR/SWIR to detect camouflaged targets. Our realtime fusion processor could easily be scaled up to support four-band fusion. Double-opponent architectures are also capable of fusing imagery of different modalities such as EO, IR, and SAR, following registration to a common reference frame. The result is a very literal image which displays each modality in the context of the others. *Such fusion architectures match opponent-sensor contrasts to human opponent-color pathways.*

REFERENCES

Barbur, J.L., & Forsyth, P.M., (1990). The effective contrast of coloured targets and its relation to visual search. Chapter 33 in **Visual Search** (ed. D. Brogan), pp. 319-328, London: Taylor & Francis.

Burt, P.J., & Kolczynski, R.J. (1993). Enhanced image capture through fusion. *Fourth Internat'l Conf. on Computer Vision*, pp. 173-182. Los Alamitos: IEEE Computer Society Press.

Ellias, S.A., & Grossberg, S. (1975). Pattern formation, contrast control, and oscillations in the short memory of shunting on-center off-surround networks. *Biological Cybernetics*, **20**, 69-98.

Fink, W. (1976). Image coloration as an interpretation aid. *Proceed. of SPIE/OSA Conf. on Image Processing*, Vol. **74**, 209-215.

Flannery, R.E., & Miller, J.E. (1992). Status of uncooled infrared imagers. *Infrared Imaging Systems*, **SPIE-1689**, 379-395.

Gouras, P. (1991). Color vision. Chapter 31 in **Principles of Neural Science** (E.R. Kandel, J.H. Schwartz & T.M. Jessell, editors), pp. 467-480, New York: Elsevier Science Publishers.

Gove, A.N., Cunningham, R.K., & Waxman, A.M. (1996). Opponent-color visual processing applied to multispectral infrared imagery. *Proceedings of the 1996 Meeting of the IRIS Specialty Group on Passive Sensors*, Vol. II, pp. 247-262. Ann Arbor: Infrared Information Analysis Center.

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, **52**, 217-257.

Grossberg, S. (1988a). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, **1**, 17-61.

Grossberg, S. (1988b). **Neural Networks and Natural Intelligence**, Chapters 1-4, Cambridge, MA: MIT Press.

Grossberg, S., Mingolla, E., & Ross, W.D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review*, **101**, 470-489.

Kaiser, P.K., & Boynton, R.M. (1996). **Human Color Vision (2nd edition)**, Washington, D.C.: Optical Society of America.

King, A.J. (1990). The integration of visual and auditory spatial information in the brain. In **Higher Order Sensory Processing** (D.M. Guthrie, editor), pp. 75-113, Manchester University Press.

Land, E.H. (1959a). Color vision and the natural image. Part I. *Proceedings of the National Academy of Sciences*, **45**, 115-129.

Land, E.H. (1959b). Experiments in color vision. *Scientific American*, (May) 84-99.

Land, E.H. (1983). Recent advances in retinex theory and some implications for cortical computations: Color vision and the natural image. *Proceedings of the National Academy of Sciences USA*, **80**, 5163-5169.

Newman, E.A. & Hartline, P.H. (1981). Integration of visual and infrared information in bimodal neurons of the rattlesnake optic tectum. *Science*, **213**, 789-791.

Newman, E.A. & Hartline, P.H. (1982). The infrared vision of snakes. *Scientific American*, **246** (March), 116-127.

Reich, R.K., Burke, B.E., McGonagle, W.M., Craig, D.M., Waxman, A.M., Savoye, E.D., & Kosicki, B.B. (1998). Low-light-level 640x480 CCD camera for night vision application. *Proceedings of the 1998 Meeting of the IRIS Specialty Group on Passive Sensors*. Ann Arbor: Infrared Information Analysis Center.

Ryan, D., & Tinkler, R. (1995). Night pilotage assessment of image fusion. *Helmet- and Head-Mounted Displays and Symbology Design Requirements II*, **SPIE-2465**, 50-67.

Savoye, E.D., Waxman, A.M., Buss, J., Hawkins, H., & Campana, S.B. (1996a). Charge Coupled Devices and Visible/IR Fusion for Night Vision. *Proceeds. of Night Vision '96*. London, August 1996.

Savoye, E.D., Waxman, A.M., Reich, R.K., Burke, B.E., Gregory, J., McGonagle, W.H., Loomis, A.H., Kosicki, B.B., Mountain, R.W., Gove, A.N., Fay, D.A., & Carrick, J.E. (1996b). Low-light-level imaging and image processing. *U.S. patent pending* (application submitted 15 April 1996); rights assigned to the Massachusetts Institute of Technology.

Schiller, P. (1992). The ON and OFF channels of the visual system. *Trends in Neuroscience*, **TINS-15**, 86-92.

Schiller, P., & Logothetis, N.K. (1990). The color-opponent and broad-band channels of the primate visual system. *Trends in Neuroscience*, **TINS-13**, 392-398.

Steele, P.M., & Perconti, P. (1997). Part task investigation of multispectral image fusion using gray scale and synthetic color night vision sensor imagery for helicopter pilotage. *Targets and Backgrounds: Characterization and Representation III*, **SPIE-3062**, 88-100.

Toet, A. (1990). Hierarchical image fusion. *Machine Vision and Applications*, **3**, 1-11.

Toet, A. (1992). Multiscale contrast enhancement with applications to image fusion. *Optical Engineering*, **31**, 1026-1031.

Toet, A., van Ruyven, L.J., & Valeton, J.M. (1989). Merging thermal and visual images by a contrast pyramid. *Optical Engineering*, **28**, 789-792.

Toet, A., & Walraven, J. (1996) New false color mapping for image fusion. *Optical Engineering*, **35**, 650-658.

Toet, A., IJspeert, J.K., Waxman, A.M., & Aguilar, M. (1997). Fusion of visible and thermal imagery improves situational awareness. *Proceeds. of the SPIE Conf. on Enhanced and Synthetic Vision 1997*, **SPIE-3088**, 177-188.

Waxman, A.M., Fay, D.A., Gove, A.N., Seibert, M.C., Racamato, J.P., Carrick, J.E., & Savoye, E.D. (1995a). Color night vision: Fusion of intensified visible and thermal IR imagery. *Synthetic Vision for Vehicle Guidance and Control*, **SPIE-2463**, 58-68.

Waxman, A.M. (1995b). Color night vision using visible/LWIR fusion (abstract). *1995 Conference on Automotive Night Vision/ Enhanced Driving*, ERIM: Ann Arbor, MI.

Waxman, A.M., Seibert, M.C., Gove, A.N., Fay, D.A., Bernardon, A.M., Lazott, C., Steele, W.R., & Cunningham, R.K. (1995c). Neural processing of targets in visible, multispectral IR and SAR imagery. *Neural Networks*, **8**, 1029-1051 (Special issue on Automatic Target Recognition; S. Grossberg, H. Hawkins & A.M. Waxman, editors).

Waxman, A.M., Fay, D.A., Gove, A.N., Seibert, M.C., & Racamato, J.P. (1996a). Method and apparatus for generating a synthetic image by the fusion of signals representative of different views of the same scene. *U.S. Patent 5,555,324*, issued 10 September 1996 (filed 1 November 1994); rights assigned to the Massachusetts Institute of Technology.

Waxman, A.M., Gove, A.N., Fay, D.A., & Carrick, J.E. (1996b). Realtime adaptive digital image processing for dynamic range remapping of imagery including low-light-level visible imagery. *U.S. patent pending* (application submitted 5 September 1996); rights assigned to the Massachusetts Institute of Technology.

Waxman, A.M., Gove, A.N., Fay, D.A., Racamato, J.P., Carrick, J.E., Seibert, M.C., Savoye, E.D., Burke, B.E., Reich, R.K., McGonagle, W.H., & Craig, D.M. (1996c). Solid state color night vision: Fusion of low-light visible and thermal IR imagery. *Proceedings of the 1996 Meeting of the IRIS Specialty Group on Passive Sensors*, Vol. II, 263-280. Ann Arbor: Infrared Information Analysis Center, ERIM.

Waxman, A.M., Gove, A.N., Seibert, M.C., Fay, D.A., Carrick, J.E., Racamato, J.P., Savoye, E.D., Burke, B.E., Reich, R.K., McGonagle, W.H., & Craig, D.M. (1996d). Progress on Color Night Vision: Visible/IR Fusion, Perception & Search, and Low-Light CCD Imaging. *Proceeds. of the SPIE Conf. on Enhanced and Synthetic Vision 1996*, **SPIE-2736**, 96-107.

Waxman, A.M., Carrick, J.E., Fay, D.A., Racamato, J.P., Aguilar, M., & Savoye, E.D. (1996e). Electronic imaging aids for night driving: Low-light CCD, thermal IR, and color fused visible/IR. *Proceeds. of the SPIE Conf. on Transportation Sensors and Controls*, **SPIE-2902**.

Waxman, A.M., Gove, A.N., Fay, D.A., Racamato, J.P., Carrick, J.E., Seibert, M.C., & Savoye, E.D. (1997). Color night vision: Opponent processing in the fusion of visible and IR imagery. *Neural Networks*, **10**, 1-6.

Widdel, H., & Pfendler, C. (1990). Visual detection performance in a vigilance task when using a colour display. Chapter 34 in **Visual Search** (D. Brogan, editor), pp. 331-338, London: Taylor & Francis.

Wolfe, J.M., Cave, K.R., & Franzel, S.L. (1989). Guided search: An alternative to the feature integration model of visual search. *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 419-433.